ChinaXiv合作期刊

知识管理论坛 ISSN 2095-5472 CN11-6036/C
Knowledge Management Forum

E-mail: kmf@mail.las.ac.cn
http://www.kmf.ac.cn

【工具方法】

基于 ISLI 标准的专利和专利参考文献关联数据集研究

朱江 [1,2] 张蒂 [1,2] 刘春江 [1,2] 周奇 [3]
1. 中国科学院成都文献情报中心 成都 610041
2. 中国科学院大学经济与管理学院图书情报与档案管理系 北京 100190
3. EBSCO International, Inc. 北京 100026

摘要：[目的 / 意义] 利用 ISLI 标准建立专利与专利参考文献的关联数据集，促进文献计量研究和科学、技术的关联研究。[方法 / 过程] 在解析专利与专利参考文献的关联方案、分析目标文献唯一标识符及其获取方法的基础上，探讨利用 ISLI 标准进行专利与专利参考文献关联数据集建设的方法。[结果 / 结论] 专利与专利参考文献关联数据集可用来丰富和完善现有的引文体系，打通专利和非专利文献两大体系，更加准确地计量专利、非专利文献的被引次数和影响力。

关键词：国际标准关联标识符　专利　非专利参考文献　关联数据集　ISLI

分类号：G251

引用格式：朱江 , 张蒂 , 刘春江 , 等 . 基于 ISLI 标准的专利和专利参考文献关联数据集研究 [J/OL]. 知识管理论坛 , 2022, 7(2): 209-217[引用日期]. http://www.kmf.ac.cn/p/287/.

1 引言

专利是集技术、经济、法律信息为一体的 [1] 信息资源，是知识产权的重要形式之一。国内外比较著名的商业性专利数据库有智慧芽（PatSnap）、incoPat、DII、DI、Innography、Orbit 等，非商业性专利数据库有欧洲专利局专利文献数据库（esp@cenet）等，其中大部分专利数据库的专利文献收录量都超过了 1.2 亿条。而非专利文献的数量则更为庞大，Summon、EDS 等发现系统可提供的各类元数据都是数十亿条。

非专利文献、专利文献分别代表了科学研究、技术开发的成果，两者之间存在着密切的关联。这种关联不仅体现在机构、人员可能是相同的，也体现在研究主题、技术方案可能存在传承、相关，更体现在参考文献存在相互引证等多个方面。

专利参考文献是指在专利文件中列出的与本专利申请相关的其他文献，第一类是专利类参考文献，第二类是科技期刊论文、著

基金项目：本文系中国科学院文献情报能力建设专项“知识资源中心体系建设”（项目编码：292020000220）和科技部国家重点研发计划重点专项“专业内容知识服务众智平台与应用示范”（项目编号：2017YFB1402400）研究成果之一。
作者简介：朱江，研究馆员，硕士，硕士生导师，E-mail: zhuj@clas.ac.cn；张蒂，硕士研究生；刘春江，副研究馆员，博士研究生；周奇，工程师，博士。
收稿日期：2022-01-26　　发表日期：2022-04-07　　本文责任编辑：刘远颖

作、会议论文等非专利参考文献[2]（Non-Patent References, NPR）。

很多学者将非专利文献和专利文献分别视作科学和技术的代表，通过专利和非专利文献中机构、人员、主题和参考文献等的关联来探索科学和技术的关联[3-4]、演化[5]和趋势[6]等。因而，如能在专利数据库中对每件专利引证的专利类和非专利参考文献实现准确的标注，并提供全文链接，将大大方便用户阅读专利说明书、理解专利的新颖性和创造性，也更有利于学者通过全文分析提升科学、技术的关联研究。

2 专利与专利参考文献的关联现状

2.1 现状分析

由于专利和非专利文献存在密切的关系，科技工作者非常希望实现专利和非专利文献的整合。目前，比较有认知度的专利和非专利文献整合系统包括 Web of Knowledge、InnovationQ Plus 等。

Web of Knowledge 平台上的 DII 数据库收录专利文献数据约 1.2 亿条，而 Web of Science 仅收录顶尖的非专利文献，其中 SCIE 收录科技类期刊 9 000 多种，科技会议录引文索引每年收录会议约 3 000 个，规模、数量偏小。Web of Knowledge 平台虽然实现了专利和非专利文献的统一检索，但检索结果集中的专利和非专利文献之间缺乏关联。

InnovationQ Plus 号称“专利与非专利文献统一检索系统”，但主要为电子、通信领域等学科服务，并非科学技术全学科。

其他众多的专利数据库虽然收录了专利参考文献，除了专利类参考文献可以通过专利号、申请号十分方便地链接外，非专利参考文献很难直接链接题录摘要信息，更不用说全文了。这一缺陷给用户查阅、链接非专利参考文献带来了诸多不便。因而有必要开展专利与非专利参考文献关联的研究，帮助用户快速、准确地定位、链接到专利引证的非专利文献的题录摘要信息，甚至全文，用户如有访问和使用权，则可直接访问和下载全文。

2.2 专利参考文献的类型

如上所述，专利参考文献包括专利和非专利两大类，以欧洲专利局的 docdb 数据为例，其非专利参考文献包括论文、图书、化学摘要、数据库、生物摘要、期刊和 Web 网站等多种类型，表 1 展示了欧洲专利局 docdb 数据中的非专利参考文献的代码、类型和示例。

表 1 欧洲专利局 docdb 数据非专利参考文献表

参考文献代码	参考文献类型	示例	
		公开号	非专利参考文献实例
A	Abstract citation of no specific kind	US10683245B2	Wang Y, Tanko J M. Radical ion probes. Part 10. Ceric (IV) ammonium nitrate oxidation of cyclopropylarenes. Journal of the Chemical Society, Perkin Transactions 2, 1998 (12): 2705-2712.
B	Book citation	US10683245B2	Wiley R H, Jarboe C H. Chemistry of heterocyclic compounds, Volume 22. John Wiley & Sons, Ltd, 1967.
C	Chemical abstracts citation	US10683245B2	Kost A N, Ershov V V. Reactions of hydrazine derivatives. III. 3-Arylpyrazolines, 1955.
E	Database citation	EP1944322B1	D’Haens G, Swijsen C, Noman M, et al. Etanercept in the treatment of active refractory Crohn’s disease: a single-center pilot trial. https://pubmed.ncbi.nlm.nih.gov/11569676/.
I	Bilogical abstract citation	EP1944322B1	Keystone E, et al. The fully human anti-tnf monoclonal antibody, adalimumab (D2E7), dose ranging study: The 24-week clinical results in patients with active RA on methotrexate therapy (the armada trial). https://ard.bmj.com/content/60/Suppl_1/A481.1.
S	Serial / Journal / Periodical citation	US10683245B2	Shabarov Yu S, Vasil’ev N I, Mamaeva N K, et al.Cyclopropanes and cyclobutanes. XXX. Cyclopropanes and cyclobutanes with p-biphenylyl and naphthyl radicals. Journal of general chemistry USSR,1963,33(7): 2199 – 2123.
W	World Wide Web / Internet search citation	EP1944322B1	REMICADE® (INFLIXIMAB) PRODUCT INFORMATION. http://www.drugbank.ca/system/fda_labels/DB00065.pdf?1265922797.

知识管理论坛，2022（2）:209-217
DOI: 10.13266/j.issn.2095-5472.2022.018

其中，代码为A（abstract citation of no specific kind）的非专利参考文献所占比例最多，包括期刊论文、学位论文和技术报告等。以美国专利US9622820B2为例，表2展示了该专利的基本信息和部分非专利参考文献信息。

表2 欧洲专利局docdb数据中美国专利US9622820B2信息实例

类型	实例
公开号	US9796668B2
专利名称	Feature-driven rule-based framework for orthopedic surgical planning
申请号	US201313871675
公开时间	2017-04-18
申请时间	2013-04-26
申请人	BALOCH SAJJAD HUSSAIN等
参考的期刊文献	Ye X, Liu H, Chen L, et al. Reverse innovative design—an integrated product design methodology. Computer-aided design, 2008, 40(7): 812-827.
参考的技术报告	Xiong G, et al. Automatic feature detection for implant design in knee replacement surgery. Siemens corporate research,2012.
参考的会议论文	Liu D, Zhou K S, Bernhardt D, et al. Search strategies for multiple landmark detection by submodular maximization.2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2010: 2831-2838.
参考的其它文献	International search report and written opinion dated Sep. 20, 2013 for corresponding PCT Patent Application No. PCT/US2013/038607.

3 基于ISLI标准的专利与专利参考文献关联数据集设计

如上所述，专利与专利参考文献的关联，本质上就是两种文献的关联，要建立专利与专利参考文献的关联数据集并实现共享，可以采用多种方法，其中一种方法是利用我国主导制定、2015年正式发布的《ISO 17316:2015 信息与文献——国际标准关联标识符（ISLI）》[7]来建立。

ISLI是一项全新理念的标识符，并不标识一个单一的实体对象，而是标识两个实体之间的关联关系[8]，其功能就是在具有特定标识符的“源”和“目标”两个实体之间建立起关联，而且这两个实体的粒度可以相同，也可以不同。ISLI的这种特点，为不同类型实体的关联提供了解决方案[9]。

3.1 ISLI关联编码方案

利用ISLI关联编码，可以将带有专利参考文献的一件专利作为“源”，将其引用的每篇参考文献作为“目标”，并赋予一个ISLI编码，就可建立起一条ISLI关联记录，其中“源”可用公开（公告）号作为标识符，“目标”可能是期刊（会议）论文、学位论文、专著、专利等多种类型，因而可用DOI、handle、ISBN-A、URN和公开（公告）号作为标识符（见图1）。

由于ISLI标志码由十进制数字构成，分为服务字段、关联字段和校验字段3个部分[10]，一件专利与专利参考文献关联的ISLI标志码可以按如下方式设计（见图2）。

第一部分为服务字段，编码一般为6位，由ISLI注册中心（ISLI RA）分配。

第二部分为关联字段，编码长度可变，且一般可细分为前置编码和后置编码两部分。根据目前全球专利量和今后的增长趋势，可将前置编码设为10位，在不扩容的情况下，可对100亿条专利进行关联编码；后置编码设为5位，除00000编码指向本专利的详细记录外，其余



图 1 基于 ISLI 标准的专利与专利参考文献关联模型



图 2 ISLI 标志码字段结构

00001-99999 编码可为每件专利设置 99 999 个专利参考文献或其它属性的关联。

第三部分校验字段的数值（校验码）由 ISLI 系统依据规则自动计算。

3.2 专利参考文献标识符的获取

3.2.1 专利类参考文献标识符的获取

在一个专利数据库系统里，专利类参考文献可以通过公开（公告）号或专利号、申请号等号码很方便地定位、链接，甚至通过一定的转换规则，可直接链接到欧洲专利局等专利网站上查阅该专利的详细记录和专利说明书全文。这种关联是两件专利或专利申请之间的关联，将两件专利的公开（公告）号作为标识符，两件专利就可以直接关联了，相对简单。

3.2.2 非专利参考文献标识符的获取

由于非专利文献类型多样，要实现专利与非专利参考文献关联则复杂很多。通过分析发现，在所有的非专利参考文献中，期刊论文是最主要的非专利文献，因此笔者以期刊论文为例，详细说明非专利参考文献标识符的获取方法。

期刊论文是非专利参考文献的主体，且出版规范，普遍实现了数字化，在网络上几乎都可查到其电子版，对于开放获取出版的期刊论文，其全文可被直接获取，对于商业出版的期刊论文，如订购了使用权，用户可在授权范围 IP 内或通过用户名密码登录后直接查看全文，如未订购使用权，一般可免费查看摘要。

以表 2 中的期刊论文“Ye X, Liu H, Chen L, et al. Reverse innovative design—an integrated product design methodology. Computer-aided design, 2008, 40(7): 812-827.” 为例，首先登录 http://www.crossref.org/guestquery/，输入期刊论文第一责任者的姓氏和论文题名，可查询到该文的永久链接 (Persistent Link)——http://dx.doi.org/10.1016/j.cad.2007.07.006，并可

知识管理论坛，2022（2）:209-217
DOI: 10.13266/j.issn.2095-5472.2022.018

获取该论文的DOI号（见图3），直接将该永久链接和DOI号写入数据库。点击该永久链接，即可跳转到出版商网站上的该篇论文（见图4），直接查看摘要等详细信息，拥有访问权的用户还可下载查看全文。DOI号则可用来标识该篇论文。



图3 期刊论文永久链接查询界面

图4 通过永久链接直接查看期刊论文摘要、获取全文示例

如无法通过上述方法获取期刊论文的永久链接和DOI号，可以利用资源发现系统来查询，下面以EDS发现系统为例进行说明。

EDS发现系统提供基于RESTful API技术的数据检索接口，返回XML/JSON格式检索结果数据，在数据加工和系统实时运行过程中可利用EDSAPI接口，通过输入“论文题名＋刊名＋日期”等复合条件，从EDS庞大的元数据仓储中获得该论文的详细信息。

以期刊论文 *Land use change moritoring in nature reserves base on GF-1/GF-2* 为例，EDS返回的JSON数据经解析后（见图5）可获得该论文的URL和其它元数据，其中系统存取号（AN）可作为该论文的标识符。

```
⊟{
    "ResultId":17,
    "Header":⊟{
        "DbId":"edseee",
        "DbLabel":"IEEE Xplore Digital Library",
        "An":"edseee.9362542",
        "RelevancyScore":"2148",
        "PubType":"Conference",
        "PubTypeId":"conference"
    },
    "PLink":"https:\/\/search.ebscohost.com\/login.aspx?direct=true&site=eds-live&db=edseee&AN=edseee.9362542",
    "FullText":⊟{
        "Text":⊟{
            "Availability":"0"
        },
        "CustomLinks":⊟[
            ⊟{
                "Url":"http:\/\/ieeexplore.ieee.org\/xpl\/articleDetails.jsp?arnumber=9362542",
                "Name":"IEEE Xplore (All Content) - (s1625991)",
                "Category":"fullText",
                "Text":"",
                "Icon":"http:\/\/imageserver.ebscohost.com\/branding\/ieee\/ieeecustomlink.png"
            }
        ]
    },
    "Items":⊟[
        ⊟{
            "Name":"Title",
            "Label":"Title",
            "Group":"Ti",
            "Data":"Land Use Change Moritoring in <highlight>Nature<\/highlight> Reserves Base on GF-1\/GF-2"
        },
        ⊟{
            "Name":"Author",
            "Label":"Authors",
            "Group":"Au",
            "Data":"<searchLink fieldCode="AR" term="%22Huang%2C+Mutao%22">Huang, Mutao<\/searchLink><br \/>
<searchLink fieldCode="AR" term="%22Wang%2C+Jinmeng%22">Wang, Jinmeng<\/searchLink><br \/>
<searchLink fieldCode="AR" term="%22Li%2C+Xiaojuan%22">Li, Xiaojuan<\/searchLink>"
        },
        ⊟{
            "Name":"TitleSource",
            "Label":"Source",
            "Group":"Src",
            "Data":"2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA) Power Ele
ctronics, Computer Applications (ICPECA), 2021 IEEE International Conference on. :228-231 Jan, 2021"
```


图 5 EDS 发现系统元数据截图（部分）

学位论文、图书专著等类型的非专利参考文献可通过类似方法获取其 handle、ISBN-A、URN 或 EDS 发现系统的 AN 号，并将其作为标识符。有了标识符后，非专利参考文献就可与专利通过国际标准关联标识符进行关联。

3.3 元数据集扩展

除了 ISLI 标准规定的基本元数据集，为准确表示专利参考文献的文献类型和标识符类型，需要设计和建立专利和专利参考文献关联数据集的扩展元数据集，对“目标”文献的文献类型、文献标识符类型以及关联方式进行代码化标识，部分扩展元数据可自动生成，无法自动生成的扩展元数据可通过多种途径或利用 EDSAPI 从 EDS 发现系统中获取并填充。

4 基于 ISLI 标准的专利与专利参考文献关联数据集的建设

4.1 专利大数据服务平台的开发

中国科学院成都文献情报中心利用欧洲专利局（EPO）XML 格式的专利裸数据，自建了专利数据服务平台（http://tmcloud.casip.ac.cn/pbsp/）（见图 6），收录了 100 多个国家、地区和组织的 1.2 亿条专利题录数据和法律状态数据，并通过大数据分析引擎构建，重点实现了智能检索、专业分析、知识发现和数据下载四大功能。其中在智能检索功能上，一方面结合

知识管理论坛，2022（2）:209-217
DOI: 10.13266/j.issn.2095-5472.2022.018

多样化的检索策略和语义检索技术实现专利检索的智能化，另一方面结合多角度的数据分面实现丰富友好的检索结果浏览模式；在专业分析功能上，结合多维度分析指标实现功能强大的分析效果；在知识发现功能上，结合大数据知识计算实现关键技术挖掘与预见；在数据下载功能上，结合自定义用户需求实现个性化下载。

4.2 资源发现系统的建设

中国科学院成都文献情报中心引进 EDS 发现系统（见图 7），对成都文献情报中心订购、链接的 100 多个商业、开放数据库的文献资源实现了一站式发现。



图 6 中国科学院成都文献情报中心自建专利数据服务平台主页



图 7 EDS 资源发现系统检索结果页面

以上两类系统分别是专利和非专利两大类型文献的集成系统，但彼此之间缺乏关联。

4.3 专利类参考文献关联的实现

在专利数据服务平台上，已通过公开(公告)号实现了专利与专利类参考文献的关联，点击专利参考文献列表中的专利类参考文献，可直接跳转到该专利的详细页面，查看该专利的详细信息。

4.4 非专利参考文献关联的实现

针对专利参考文献列表中的非专利参考文献，开发专门程序，获取期刊类参考文献的永久链接和 DOI 号，写入专利数据库并展示在参考文献列表中，用户点击该永久链接即可跳转到出版商或服务商平台上查看该期刊论文的摘要，拥有全文访问权的用户还可直接下载、查看全文；查不到永久链接或没有永久链接的期刊类参考文献则调用 EDSAPI 接口，获取 EDS 发现系统的 AN 号、全文链接地址和其它元数据信息，一并写入专利数据库，并将全文链接地址展示在参考文献列表中，用户可以点击该全文链接地址实现跳转。

4.5 专利与专利参考文献关联数据集的生成

利用 ISLI 标准生成专利与专利参考文献关联数据集，需为每一件专利申请一个 ISLI 基本标识码（如 ISLI 012345-012345678900000-8）及其附带的 99 999 个标识码（后置编码为“00001”-“99999”）。ISLI 基本标识码的“源”和“目标”都标记为该专利的公开（公告）号，再依次为每篇参考文献的关联分配一个 ISLI 标识码，“源”均标记为该专利的公开（公告）号，“目标”则依次标记为每篇参考文献的公开（公告）号、DOI、handle、ISBN-A、URN、EDS 发现系统的 AN 号等标识符，再将每个 ISLI 标识码关联的文献类型、参考文献的标识符类型、从多种途径补充的元数据等写入元数据集。

目前，中国 ISLI RA 尚未全面开通 ISLI 服务代码的申报和解析服务，各种 ISLI 标准应用系统可自建模拟服务器对 ISLI 标识码进行申领、编码和解析，待 ISLI 服务代码申报开放后再将自编的模拟 ISLI 标识码批量转换为正式的 ISLI 标识码并上传 ISLI RA。

按照此方法建立的专利与专利参考文献关联数据集不仅可以共享复用，还可以用来计量不同国家（地区、组织）、不同部类（来源地、机构、发明人）专利参考文献的总量、平均量和极值，从而分析、掌握其差异和特征。同时，还可将该数据集中的“源”和“目标”翻转，建立专利参考文献与专利“镜像”关联数据集，找到不同学科领域被专利引用次数最多的期刊论文、学位论文、会议论文、专著等，在现有的期刊论文、会议论文、图书、科学数据等引文数据库基础上新增专利引文数据库。同时，也可在非专利文献平台（如 EDS 发现系统）直接调用该“镜像”关联数据集，提供非专利文献被专利引用的数据（见图 8），供读者查询、参考。

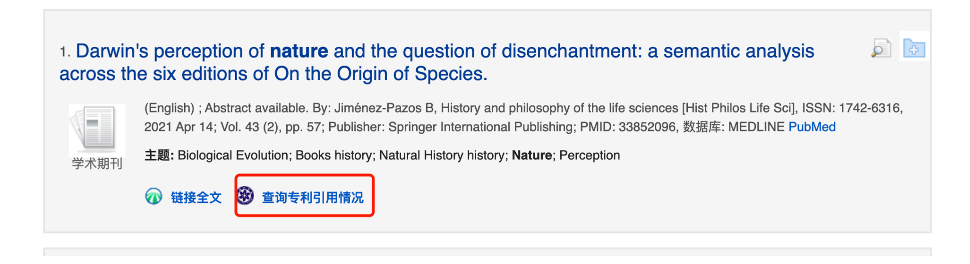


图 8 EDS 资源发现系统中的专利引用链接

5 结论

众所周知，以 Web of Science 和 CNKI 为代表的引文数据库系统目前还没有收录期刊论文、会议论文、专著被专利引用的数据，利用

该专利参考文献数据集可十分方便地增加这些引文数据库中的收录文献被专利引用的数据，从而进一步丰富和完善现有的引文体系，更加准确地计量收录文献的被引次数和学术影响力，并为基础研究成果和技术创新成果的演化、推进研究提供引文数据支撑。但由于很多早期的非专利文献目前还缺乏 DOI 这样规范、统一的标识符，目前只能暂时借用 EDS 发现系统的 AN 号来标识这些早期的非专利文献，由于非 EDS 发现系统用户无法解析这些 AN 号并链接到 EDS 发现系统的相关记录，导致专利与专利参考文献关联数据集的可用性有所降低。随着 DOI 等权威标识符的不断回溯，使用 EDS 发现系统 AN 号作为标识的比例将会逐渐减少，专利与专利参考文献关联数据集的可用性将会得到逐步提升。

参考文献：

[1] 陈凯 , 徐峰 , 程如烟 . 非专利引文分析研究进展 [J]. 图书情报工作 , 2015, 59(5): 137-144.

[2] 李时玉 , 郭建伟 , 孙沫卿 . 专利检索技巧与方法之引证检索 : 以奥凯知识产权数据资源为例 [J]. 河南科技 , 2018(6): 59-61.

[3] NARIN F, NOMA E. Is technology becoming science?[J]. Scientometrics, 1985, 7(3-6): 369-381.

[4] GUAN J, HE Y. Patent-bibliometric analysis on the Chinese science-technology linkages[J].Scientometrics, 2007, 72(3): 403-425.

[5] 韩芳 . 基于专利引文的“科学—技术关系”及技术演化轨迹研究 [D]. 北京 : 北京邮电大学 , 2017.

[6] MEYER M S. Patent citation analysis in a novel field of technology: an exploration of nano-science and nano-technology[J].Scientometrics, 2001, 51(1): 163-183.

[7] ISO 17316: 2015 Information and documentation—International standard link identifier (ISLI)[EB/OL]. [2021-09-16].https://www.iso.org/standard/59560.html.

[8] 蔡逊 . 关联 :ISLI 国际标准的理念与价值 [J]. 出版发行研究 , 2015(7): 39-41.

[9] 朱江 , 李欣怡 , 任晓亚 , 等 . 国际标准关联标识符及其在图书馆的应用研究 [J]. 图书馆建设 , 2020(2): 56-61, 69.

[10] ISLI 应用指引（信息内容产业）[EB/OL]. [2021-09-16]. https://www.isli-international.org/download/Guide_for_using_ISLI.pdf.

作者贡献说明：

朱　江：提出研究思路，撰写论文；
张　蒂：论文修改和补充；
刘春江：专利大数据服务平台开发，专利参考文献类型和实例总结；
周　奇：资源发现系统关联方案设计实例的实现。

Study on Linkage Data-Sets of Patents and Patent References Based on ISLI Standard

Zhu Jiang[1,2] Zhang Di[1,2] Liu Chunjiang[1,2] Zhou Qi[3]
[1]Chengdu Library and Information Center, Chinese Academy of Sciences, Chengdu 610041
[2]Department of Library, Information and Archives Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190
[3]EBSCO International, Inc., Beijing 100026

Abstract: [Purpose/Significance] This paper aims at establishing the linkage data-sets of patents and patent references based on ISLI standard, and promoting bibliometric research and science and technology linkage research. **[Method/Process]** On the basis of analyzing the linkage schemes of patents and patent references, analyzing the unique identifiers of target documents and their acquisition methods, the method of constructing the linkage data-sets of patents and patent references by using ISLI standard was discussed. **[Result/Conclusion]** The linkage data-sets of patents and patent references can be used to enrich and improve the existing citation systems, break through the two systems of patents and non-patent literatures, and measure the number of citations and influence of patents and non-patent literatures more accurately.

Keywords: international standard link identifier　patent　non-patent references　linkage data-set　ISLI